



# Evaluating the capacity of clinical pathways through discrete-event simulation

Brecht Cardoen and Erik Demeulemeester

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

# Evaluating the capacity of clinical pathways through discrete-event simulation

Brecht Cardoen, Erik Demeulemeester

Katholieke Universiteit Leuven, Faculty of Economics and Applied Economics, Department of Decision Sciences and  
Information Management, Naamsestraat 69 B-3000 Leuven, Belgium, brecht.cardoen@econ.kuleuven.be,  
erik.demeulemeester@econ.kuleuven.be

## Abstract

Organizing a medical facility efficiently is hard due to the numerous patient trajectories and their use of joint and scarce resources. Moreover, these trajectories tend to be complex and characterized by uncertain medical processes. In this paper, we will structure patient trajectories using clinical pathways and aggregate them in a discrete-event simulation model. This model enables the health manager to evaluate and improve important performance indicators, both for the patient and the hospital, by conducting a detailed sensitivity analysis. Two case studies, performed at large hospitals in Antwerp and Leuven (Belgium), will be introduced and briefly discussed in order to illustrate the generic nature of the model.

*Keywords:* health care operations, discrete-event simulation, capacity management, case studies

---

## 1 Introduction

Increased competition leads health care institutions to rigorously monitor the costs associated with their medical services. Although a reduction in these costs, for instance through improved resource utilization, is beneficial for the health care institution, a price might be paid in terms of patient satisfaction due to the increased waiting times. Since this service concept can be exploited to become a competitive advantage, it should not be neglected and consequently thoroughly investigated (Visser, Adan and Dellaert 2006). In hospitals, however, balancing the operational costs and the service level is hard to do due to the complex structure of its constituent and interrelated medical facilities or departments. Moreover, the stochastic nature of the medical processes triggers uncertainty in the system. This implies that a flexible technique is needed in order to gain structural insights. One such technique, that already has been proven to be successful in many health care applications, is discrete-event simulation.

A diversity of simulation studies can be found in the literature. Many of these studies deal

with the operating theater utilization, on the one hand, and the hospital bed requirements, on the other hand. Bowers and Mould (2005), for instance, investigate to which degree the operating theater utilization is influenced by changing the proportion of orthopaedic day-care surgeries and inpatient treatments. In another study, they examined a policy to include elective patients within the trauma sessions (Bowers and Mould 2004). The demand for beds in the intensive care unit and the related bed-reservation schemes were the focus of research by Ridge et al. (1998) and Kim et al. (2000). Other health care applications focus, for instance, on appointment scheduling (Klassen and Rohleder 1996, Rohleder and Klassen 2000, Harper and Gamlin 2003) or the laboratory practice (Couchman, Jones and Griffiths 2002). We refer to Jun, Jacobson and Swisher (1999) for a detailed literature review on the use of discrete-event simulation for health care applications.

Although the literature on health care simulation applications in hospitals is vast, only few studies describe the entire trajectory patients have to fulfil. Swisher et al. (2001), for instance, built a detailed simulation model of an outpatient family practice clinic. This paper, though, focuses on a generic hospital setting. In particular, we will use the operational facets of *clinical pathways* to structure the simulation model. Clinical pathways, also referred to as care pathways, critical pathways, integrated care pathways or care maps, were introduced in 1985 at the New England Medical Center (Zander 2002). They describe, both qualitatively and quantitatively, for a homogeneous patient group which medical actions should be performed by the multidisciplinary care team. The sequence and time frame of the interventions is also mentioned. This standardization of the care process should eventually lead to improved quality of care, reduced risks, increased patient satisfaction and increased efficiency in the use of resources (De Bleser et al. 2006). It should be clear from Table 1 that the concept of clinical pathways is internationally accepted and consequently adopted all over the world (Zander 2002). In the next section, we will highlight the aspects of clinical pathways that are informative for building the simulation model.

The remainder of this paper is organized as follows. The simulation model will be introduced in Section 2. We will show the generic nature of the model and describe its capabilities and internal logic both with respect to the consultation suite and the surgery suite. In Section 3, we will discuss the applied methodologies for model verification and validation. The applicability of the model in different medical settings will be illustrated in Section 4. Two case studies will

be introduced and discussed by means of a sensitivity analysis. Finally, in Section 5, conclusions will be formulated and ideas for future research will be mentioned.

## 2 Model Building

In this section, we will introduce the discrete-event simulation model and list the capabilities and options that are embedded in the application. Obviously, the options chosen to simulate a patient trajectory can differ between hospitals, doctors or clinical pathways and contribute to the generic nature of the model. The model currently allows for the simultaneous aggregation of 15 clinical pathways, 10 doctors, 5 operating rooms and 5 hospital wards, but could easily be extended. A description of both the consultation suite and the surgery suite will be provided. We implemented the model using the Arena simulation software, version 8.0 (Kelton, Sadowski and Sturrock 2004).

### 2.1 A generic simulation model

Each clinical pathway can be simplified and represented by the decision scheme depicted in Figure 1. A suite of pre-surgical consultations typically precedes the surgical process of a patient, whereas this surgical process on its turn precedes a suite of post-surgical consultations. We allow, though, deviations from this structure. We can think, for instance, of clinical pathways that solely consist of consultations, or patients for which post-surgical consultations are executed in peripheral hospitals. When patients already fulfilled a dedicated suite of post-surgical consultations, they can switch to, for instance, a standard two-yearly consultation visit. Although these two-yearly consultations can strictly be seen as post-surgical too, we will refer to them as repetitive consultations (see Section 4.1). The specific combination of surgery and consulta-

Table 1: Spread and adoption of clinical pathways over the world.

Invention	1985	New England Medical Center in Boston (United States)
First wave	1986-1990+	United States, Australia, United Kingdom
Second wave	mid 1990's	Spain, New Zealand, South Africa, Saudi Arabia
Third wave	late 1990's	Belgium, Japan, Singapore, Germany
Fourth wave	2000+	Korea, Ecuador

tions for each clinical pathway will constitute the spine of the patient-driven simulation model. Although the simulation model was initially constructed to simulate clinical pathways entirely (see Section 4.1), attention can easily be focused and limited to a subsection of the pathways (see Section 4.2).

With the development of the simulation model, we aim at the evaluation of both the patient flow and the resource capacity. This analysis, however, can be performed at different levels. On the one hand, there is a long-term perspective or macro level. This implies, for instance, that we are interested in the time between the request for surgery and the actual surgery day, the utilization of the operating theater as a whole or the number of consultations that can be planned yearly for a specific doctor. On the other hand, we also allow for a short-term perspective or micro level. In this case we are, for instance, interested in the average waiting time of a patient in the consultation waiting room or the number of surgeries executed in overtime on a specific surgery day. Both perspectives or levels are integrated in the model and can be analyzed simultaneously.

In the next two subsections, we will try to illustrate the sense of detail incorporated in the model. This, however, does not mean that information is needed for every single detail in order to run the model and get accurate results. A surgery, for instance, consists of multiple phases: induction, skin-to-skin surgery, after care and cleaning of the operating room. When data for these distinct phases are not readily available, one single data distribution can be inserted for the whole of the processes. During the conduct of the case studies (see Section 4), we fully

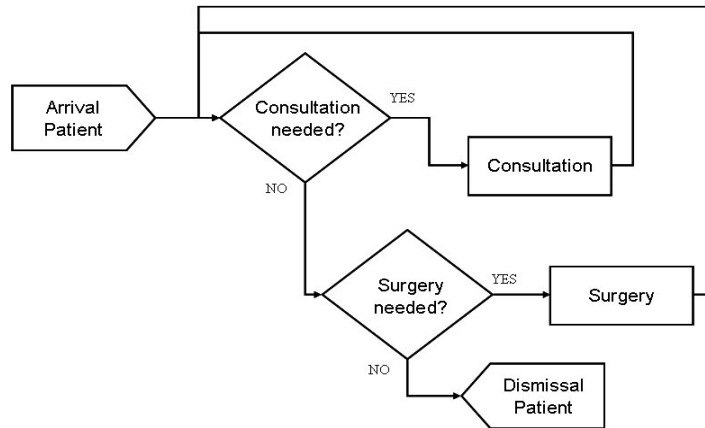


Figure 1: Decision scheme for representing clinical pathways.

experienced the practical contribution of this aggregation feature.

## 2.2 Consultation suite

The arrival pattern of patients depends on the clinical pathway to be followed. A qualified doctor is assigned to the patient for the entire clinical trajectory and an appointment for a first hospital visit is made in correspondence with the schedule of both the doctor and the patient. The amount of time reserved for a consultation may vary between the clinical pathways, the doctors and the number of previous consultations of the patient (i.e. the status of a patient). Cayirli and Veral (2003) refer to this appointment rule as the individual-block/variable-interval rule since patients are scheduled individually and appointment intervals have a variable length. Obviously, the capacity for consultation is limited and depends on the two-week, cyclic schedule of the doctor. The population of patients to be consulted by a doctor on a certain day is not structured. This means that the proportion of patient types can differ from day to day and no specific capacity is reserved for a particular clinical pathway. New patients can possibly leave the system whenever the waiting time until the consultation day is too elevated with respect to their preferences. This phenomenon, i.e. seeking alternative care in other hospitals, is referred to as *doctor shopping* and becomes increasingly important (Yeung, Leung, McGhee and Johnston 2004). When an appointment is finally planned, the patient waits at home until the day of consultation.

The sequence in which patients are seen by the doctor may vary based on the clinical pathway and the status of the patient. The arrival time of the patient in the waiting room possibly differs from the planned arrival time and there is an option to include no-shows. When the doctor is unavailable, two procedures can be initiated. First, when the doctor is busy seeing another patient, patients wait until the doctor is available. For each doctor and day of the week, it should be specified if the doctor can perform consultations in overtime due to delays. When overtime is possible, no appointments are canceled. Second, patients can opt to leave the waiting room. This occurs, for instance, when the doctor stops consulting due to an emergency case. Obviously, a new appointment has to be made as soon as possible.

During the consultation, the doctor establishes the need for possible supplementary examinations (e.g. an X-ray). If such examinations are inevitable, the consultation is interrupted and the patient leaves the consultation area to do the required tests. When the supplementary

examinations are performed and the consultation session of the doctor is not yet finished, the interrupted patient is further consulted with priority. Otherwise, a new appointment has to be made.

At the end of a consultation, a diagnosis is stated. Five possible outcomes may occur. First, the doctor can direct the patient to the help desk in order to schedule a new, future appointment. Second, it might be possible that further treatment is superfluous and that the patient prematurely leaves the clinical pathway. This happens, for instance, when patients need treatment that differs from the clinical pathway they're in. Third, the patient might have fulfilled the entire clinical pathway and should consequently leave the system. Fourth, patients need a new, future appointment, but opt to make an appointment in a peripheral hospital. They leave the system as well. Note that this fourth measure allows to incorporate the possible death of patients based, for instance, on age patterns. Finally, the doctor can urge the patient to have a surgery.

### **2.3 Surgery suite**

Patients who acknowledge the need for a surgery can either immediately make an appointment or postpone the appointment process due to professional or personal reasons. The surgical appointment system can be configured in two ways and is based on a two-week, cyclic capacity scheme. On the one hand, it is possible to introduce a planning system similar to that of the consultation suite. On the other hand, the types and the number of surgeries that might be performed on a certain surgery day can be predetermined. In this case, the surgery day is divided in specific slots and patients are only assigned to eligible slots. In other words, capacity is reserved for certain clinical pathways. The sequence in which surgeries will be performed can be determined by specifying priorities.

When a suitable surgery date is found and confirmed, patients wait at home and enter the hospital upon agreement. This day of reception does not necessarily equal the surgery date. Depending on the clinical pathway, patients may arrive at the hospital multiple days in advance for preparation. The model allows that arrivals of patients for which the reception date is equal to the surgery date are based on the expected surgery start time instead of imposing a simultaneous arrival of these patients on a joint time indication. Patients are assigned to beds in specific hospital wards when they enter the hospital. Three types of hospital wards can be

represented in the model. First, there are wards which are continuously in use and hence suited for hospitalized patients. Second, wards may daily close so that they are suited for day-care services. Finally, the simulation model allows for hospital wards that are a combination of the previous two types. Think, for instance, of a ward for hospitalized patients that is closed on Saturday and Sunday. Patterns can be specified for the transfer of patients to alternative wards when beds are unavailable in the preferred bed zone.

The surgical process itself starts with an induction phase in which the patient is anaesthetized and the operating room is prepared. Next, the actual skin-to-skin surgery is performed and after care is given to the patient. The end of the after care initiates the transfer of the patient to the recovery and consecutively to the respective ward. Note that unexpected complications can trigger a change in both the hospital ward (e.g. day-care patients that now have to be hospitalized) and the length of stay. Obviously, the operating room is thoroughly cleaned after transferral of the patient to the recovery. When this cleaning is finished, a next surgery can be started. Due to the stochastic nature of the medical processes, the actual and planned surgery start time will probably differ. In other words, the actual surgery start times highly depend on the preceding suite of surgeries. However, this is not the only factor that possibly causes deviation from the planned surgery start times. The model includes late arrivals of both patients (e.g. due to supplementary examinations) and surgeons. Moreover, emergency cases can heavily disrupt the surgery schedule. The prioritized emergency cases cause delays, so that the elective surgeries possibly have to be rescheduled or even canceled when overtime is limited or not an option. A new surgery date has to be found for the canceled and hence prioritized patients. These patients possibly have to leave the hospital since hospitalization until the new date might be too expensive. After a successful surgery and an appropriate recovery in the hospital wards, patients are dismissed. When specified in the clinical pathway, the suite of post-surgical consultations should now be initiated.

### 3 Verification and Validation

Every simulation model should be thoroughly verified and validated in order to gain valuable results. The verification process, on the one hand, deals with building the model right and is situated in the transition from the conceptual model to the computerized representation (Banks,



Carson, Nelson and Nicol 2001). The validation process, on the other hand, questions whether we are building the right model and relates to the transition from the real system to the simplified, conceptual model. Although the goal of the processes is different, they are mostly conducted simultaneously.

We applied two basic approaches in order to verify the simulation model. First, we did continuously check whether the output of the simulation runs was reasonable. This was done both for regular runs and runs in which extreme or unusual cases were tested. Second, we thought of a structured and slightly animated graphical representation of the model in which modules are briefly documented. Especially the animation did enable us to identify system anomalies. Figure 2 provides a snapshot of this graphical representation. In order to avoid unwanted modifications to the modules, users of the model can change data and options through the use of tables. In other words, they do not have to enter the decision modules in order to adapt the system to represent their situation of interest.

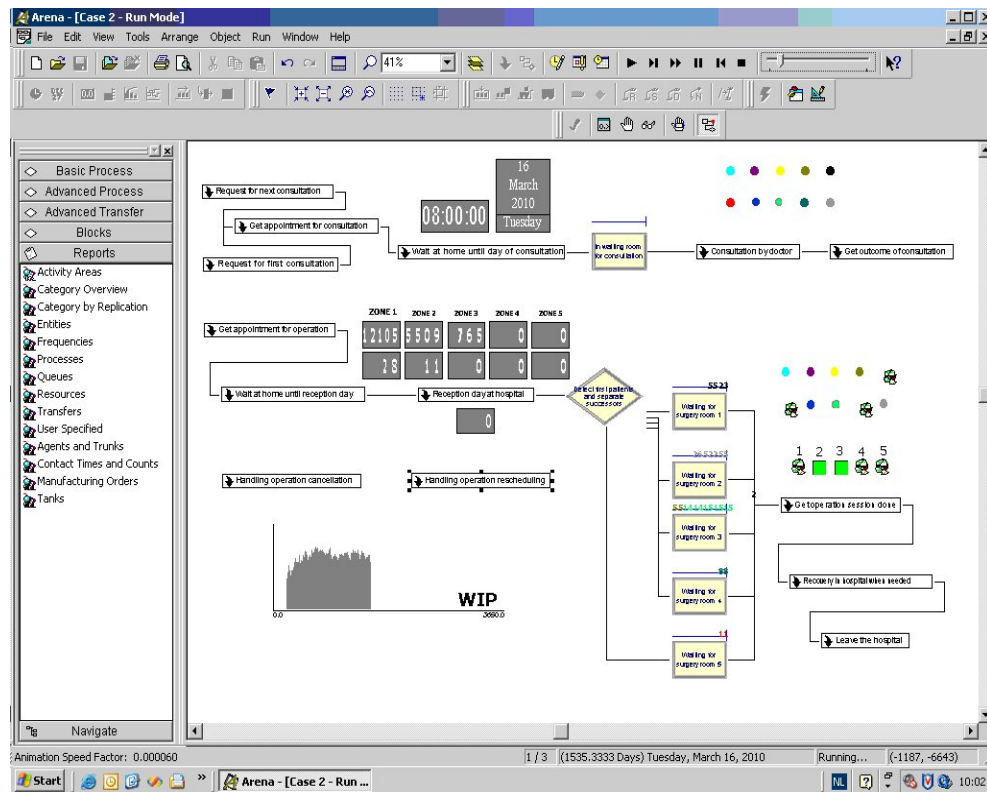


Figure 2: Graphical representation and animation of the simulation model.

During the model building and the conduct of the case studies, we considered the importance of face validity. We tried to involve people from several medical fields (doctors, clinical coordinators, etc.) and stimulated a close cooperation through meetings in which progression was discussed. However, face validity is not sufficient. Both the input and the output of the simulation model should be validated too. With respect to the input, we tried to identify probability distributions that accurately represented the registered data observations. These distributions were consequently statistically analyzed by means of goodness-of-fit tests (e.g. mean squared error,  $\chi^2$ -test or Kolmogorov-Smirnov test) and consequently accepted to be used in the simulation or rejected. In this latter case, an empirical probability distribution was introduced. With respect to the output, we mainly appealed to the daily practice of medical experts and managers to confirm the match between the reality and the simulated results.

## 4 Sensitivity Analysis

In this section, we will illustrate the operational applicability of the simulation model by means of two case studies conducted in Belgian hospitals. The first case study deals with the simulation of both the consultation and the surgery suite of multiple clinical pathways for a single doctor, whereas the second case study focuses on the surgery suite of an entire medical facility. We already mentioned in Section 2 that other settings, for instance investigating one specific pathway for multiple doctors, are equally manageable. Both case studies will be briefly situated and selected items of the conducted sensitivity analysis (i.e. alternative scenarios) will be highlighted. Although most of these scenarios are realistic in nature, health managers pointed at some potential difficulties that could arise during implementation (see Section 4.2). Note, however, that a hypothetical analysis leads to a better comprehension of the current practice too and therefore also contributes to the development of the care process. During the conduct of the case studies, thorough attention was given to the determination of the warm-up period, the replication length and the number of replications in order to get reliable and relevant outcome.

### 4.1 Case A: Orthopaedic Pathway Aggregation

The first case study was conducted at the orthopaedic facility of the Middelheim hospital in Antwerp, which is part of the Ziekenhuis Netwerk Antwerpen (6000 employees). In particular,

8 clinical pathways (e.g. total hip replacement, ligament reconstruction or unicondylar knee replacement) were specified in order to represent the entire workload of one specific orthopaedic doctor. Since the simulation of surgical processes will be extensively illustrated in the second case study (see Section 4.2), the area of interest in this case study is mainly restricted to the consultation suite and the hospital wards.

Today, the organization of the consultation practice at the orthopaedic department tends to be difficult and could be identified as a bottleneck process. The current scheduling practice (BASE), in which all time slots are equal and have a duration of 10 minutes, does not differentiate between the types of consultations and results in poor performance on mainly 3 levels (see Table 2). First, there is the elevated number of days between the request for a first consultation and the effective accomplishment of that consultation. This implies that the number of consultations that have to be performed is nearly equal to the number of time slots available. Second, the doctor has to perform a substantial amount of overtime in order to finish the scheduled program for one consultation session. This is reflected in his or her utilization, which surpasses 100%, and the percentage of consultations started or restarted during overtime. Finally, the time that is spent by patients in the waiting room, amounts to half an hour and is considered to be too large.

In section 2, we already mentioned the phenomenon of doctor shopping. Since the time span between the request for consultation and the day of consultation is large, there is a substantial risk that patients will seek alternative care. In order to avoid lost profits and to increase the patient satisfaction, this performance indicator should be improved. Multiple alternatives could now be specified in order to influence this time span. In a first alternative (A1), we will reduce the planned consultation time of the repetitive consultations from 10 minutes to 9 minutes. This way we hope to enlarge the population of patients that can be consulted during the available consultation time. Moreover, the match between the actual and the planned consultation duration should be better since repetitive consultations are often performed in less than 10 minutes. Although this strategy seems to be effective in reducing the time between the request for consultation and the accomplishment of that first consultation to 17 days, it also entails a deterioration of the other performance indicators that are represented in Table 2. Since more patients have to be consulted in one session, a delay will affect the patient satisfaction of an increased number of patients. This increased number of patients possibly also entails an increased amount of

Table 2: Performance indicators for the tested alternatives (Case A).

	BASE			A1			A2		
Doctor utilization (%)	111.16 (0.11)			112.63 (0.46)			111.50 (0.30)		
	Ba	A1	A2	Ba	A1	A2	Ba	A1	A2
	×	✓		✓	×	✓		✓	×
Time in consultation waiting room (minutes)	28.95 (0.15)			33.30 (0.23)			20.48 (0.19)		
	Ba	A1	A2	Ba	A1	A2	Ba	A1	A2
	×	✓	✓	✓	×	✓	✓	✓	×
Nr of patients in waiting room on arrival	2.22 (0.02)			2.65 (0.02)			1.34 (0.02)		
	Ba	A1	A2	Ba	A1	A2	Ba	A1	A2
	×	✓	✓	✓	×	✓	✓	✓	×
Consultations (re)started in overtime (%)	10.96 (0.09)			12.00 (0.28)			9.15 (0.13)		
	Ba	A1	A2	Ba	A1	A2	Ba	A1	A2
	×	✓	✓	✓	×	✓	✓	✓	×
Time between request and first consultation (days)	40.66 (8.01)			17.14 (5.49)			33.47 (11.83)		
	Ba	A1	A2	Ba	A1	A2	Ba	A1	A2
	×	✓		✓	×				×

✓ denotes a significant difference at a 95 percent confidence level

(x.xx) denotes the half width for a 95 percent confidence interval

variability in the execution of the consultation programme. Moreover, the number of planned minutes that were superfluous due to the overestimation of the expected duration decreased. This implies that the repetitive consultations do not have the same strength anymore to act as a buffer in order to reduce accumulated delays by preceding patients, so that most performance indicators deteriorate. Although the results are conform to the expectations, a contribution can be found in the quantification of the effects. In a second alternative (A2), we will undo the measures of the first alternative and investigate what would be the impact of eliminating both the need for supplementary examinations on the day of the consultation and the deviation from the expected arrival time of the patient in the waiting room. In other words, we will assume that all supplementary examinations are performed by, for example, the family doctor and that there

is no uncertainty in the arrival pattern of the patients in the waiting room. The aim of these measurements is to reduce the number of consultations that has to be planned. Since patients who are returning late from their supplementary examination, i.e. when the doctor has already ended his or her consultation session, have to make a new appointment, reducing the need for supplementary examinations should also reduce the need for making new appointments and hence reduce the consumption of time slots. A similar reasoning applies to the reduction in the variability of the arrival time of the patient in the waiting room. Reducing the probability that patients arrive in the waiting room when the doctor has already ended his or her consultation session should also reduce the need to make new appointments. Although these measurements tend to decrease the number of days between the request for a first consultation and the accomplishment of that consultation, this performance indicator does not significantly differ from BASE (see Table 2). However, reducing the uncertainty triggered by the supplementary examinations and the arrival pattern seems to be beneficial w.r.t. the percentage of consultations that have to be started (not restarted since the probability to interrupt the consultation is now equal to 0) in overtime. Moreover, there is a significant decrease in the number of patients in the waiting room upon arrival and the time spent by the patients in the waiting room. Note that this latter performance indicator is not equal to the consultation waiting time since patients may explicitly choose to arrive early. These early minutes cannot be seen as waiting time since it is the patient's preference to be already present in the waiting room. Although we only discussed two alternatives, it should be clear that other viable options, for instance an increase in the consultation capacity or a decrease in the doctor's workload by reducing the number of post-surgical consultations in a specific clinical pathway, could be investigated.

Although institutions often introduce structural changes to improve their operational efficiency, for instance by increasing the consultation availability of the doctors, this should not always be needed to improve performance indicators. In order to illustrate this proposition, we will use the simulation model to test the impact of patient sequencing on both the number of consultations started or restarted in overtime and the consultation waiting time. We already mentioned the discrepancy between the consultation waiting time and the time spent by patients in the waiting room. However, it should be clear that reducing the waiting time will also trigger a reduction in the time spent in the waiting room. The simulation setting is equal to that of BASE and we will differentiate between three categories of consultations. A first category (A)

captures consultations of patients who made an appointment for the first time. The second category (B) consists of the remaining pre-surgical or post-surgical consultations in the clinical pathways, whereas the third and final category (C) represents the repetitive consultations. Obviously, other categorizations of the consultations could equally be tested. The results of 13 consultation sequencing rules are depicted in Figure 3. In this figure, C-A-B represents a rule in which repetitive consultations are scheduled first on the consultation day, followed by the consultations of new patients, which are, on their turn, followed by consultations of category B. B-Rand(AC) denotes a sequencing rule in which consultations of category B are scheduled first. For the sequel of the consultation day, however, there is no explicit order between the consultations of new patients and the repetitive consultations. In order to interpret the results, we are interested in finding the efficient frontier. This efficient frontier visualizes only those sequencing rules that are efficient, i.e. rules for which no other sequencing rule could be found that has a better performance regarding the two performance indicators of interest. In Figure 3, the efficient frontier only consists of two scheduling rules, namely B-C-A and Rand(BC)-A. These rules significantly differ from each other at a 95 percent confidence level both w.r.t. the consultation waiting time and the percentage of consultations (re)started in overtime. They both indicate that it is advantageous to schedule consultations of category A near the end of the consultation session. Consultations of this category are characterized by a higher variability and often tend to take much longer than the reserved 10-minute time slot. This implies that the consultations of category A easily cause delays in the consultation programme of the doctor. Note that delays in the beginning of the session can influence the entire consultation programme, whereas delays at the end will reduce the service level of only a limited number of patients. This conclusion is in correspondence with the findings of Klassen and Rohleder (1996). They state that good results can be obtained when patients with large service (consultation) time standard deviations are scheduled toward the end of the appointment session. We already mentioned that the repetitive consultations consume less time than the capacity reserved by the time slot. This implies w.r.t. B-C-A that consultations of category C diminish the waiting time accumulated by the consultations of category B. For C-B-A, on the contrary, waiting times will typically be accumulated when the consultations of category C are already performed and hence cannot be used anymore to reduce congestion in the waiting room. The time gains accumulated in the beginning of the session cannot fully be exploited since the arrival of patients is based on a

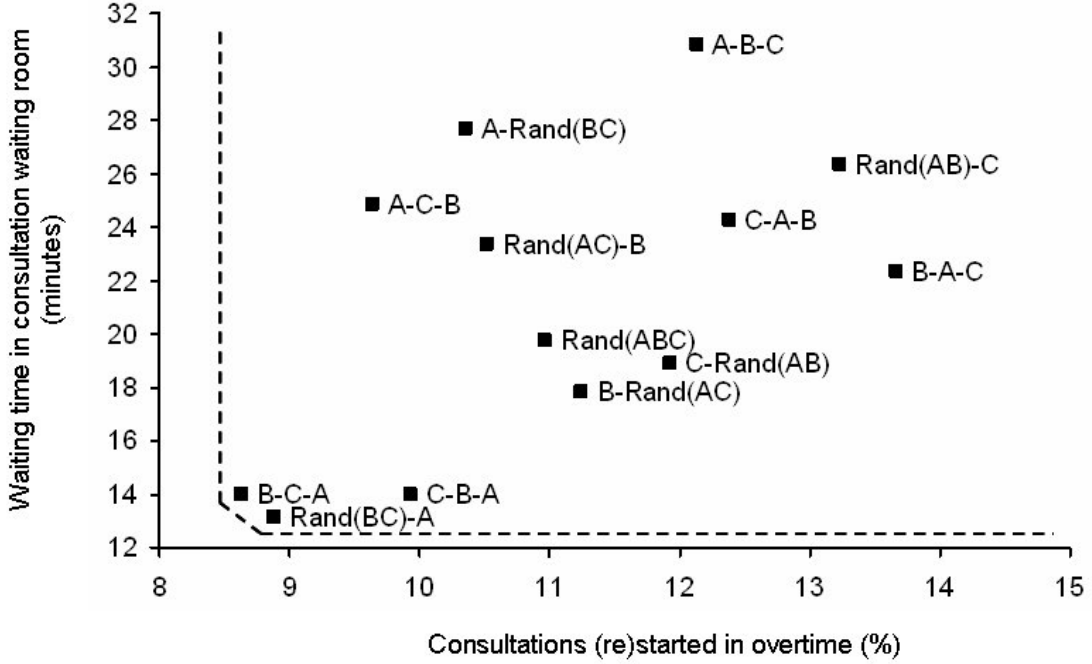


Figure 3: Efficient frontier of consultation sequencing rules w.r.t. the base model.

schedule. This results in an increased percentage of consultations that is started or restarted in overtime. A reason for the (good) positioning of C-B-A w.r.t. the consultation waiting time can be found in the discrepancy between the limited waiting time of consultations of category C and the higher waiting times of the other consultations. Rand(BC)-A is obviously a combination of C-B-A and B-C-A. Note that the interpretation of the outcome in Figure 3 is not straightforward since results are biased due to the deviation from the expected arrival time in the waiting room and the occurrence of supplementary examinations.

The utilization of the hospital wards will be the last topic to highlight in this first case study. More specifically, we will restrict the focus to the hospitalized patients. Regardless of exceptions, hospitalized patients are centralized in one hospital ward. The average number of beds occupied by patients of the doctor of interest is equal to 7.23, whereas the maximum number of beds that is simultaneously occupied reaches to 17 beds. We can now use the simulation model in order to show that patient pooling is advantageous with respect to the maximum number of beds that should be provided. In particular, we will investigate how many beds should be available when the patient population is divided over two hospital wards instead of one hospital ward. Since one clinical pathway, i.e. the arthroscopy of the knee, implies day-care treatment, we only

have to separate 7 clinical pathways over the two hospital wards. Moreover, we will only direct patients who participate in the clinical pathway of ligament reconstruction to a separate ward since this is the only pathway in which no prosthesis is used. Separating the pathways, however, results in an increased maximum need of  $16+3=19$  hospital beds. When the peak demand for beds in both hospital wards takes place at exactly the same point in time, patient pooling would not reduce the maximum number of hospital beds needed. The probability that this situation occurs, however, is small. Both peaks will hardly ever be realized simultaneously, which makes patient pooling interesting.

Since we are already pooling the hospitalized patients on one orthopaedic ward, we could argue whether it would still be possible to reduce the maximum number of beds needed or not. Since the occupation of the hospital beds somehow depends on the surgery schedule, we can try to rearrange the assignment of specific surgeries into the large operating theater blocks to be filled by the doctor. In particular, when we assume that the doctor performs revisions of total hip or knee replacements solely on a Wednesday instead of a Monday or a Thursday, the simulation model indicated that the maximum number of beds needed dropped from 17 to 16. Although this result is reached by trial and error and is not funded by an optimization tool, it clearly captures the interdependency of the hospital wards with the surgery schedule. Note, again, that the reassignment of the specific surgeries is not a structural change.

Finally, we could also question whether it is really necessary to provide 17 hospital beds for the doctor's patient population. In Figure 4, we plotted the impact of reducing the number of available beds in the orthopaedic ward on a second, alternative ward. Patients are assigned to this alternative ward when the hospital beds of the orthopaedic ward are fully occupied. From the graph, we notice that a reduction in capacity from 17 beds to 14 beds does not substantially affect the bed utilization on the alternative ward. This implies that there is only a sporadic demand of 14 beds or more so that these situations can be seen as outliers. However, this situation changes with a further restriction of the beds. The utilization of alternative hospital beds seems to increase non-linearly. In other words, reducing the number of orthopaedic beds from 13 to 12 leads to a smaller increase in the use of alternative beds than a reduction from 9 to 8 beds. Note that when we would limit the number of orthopaedic beds to the average number of beds used in the base case, i.e.  $7.23 \approx 7$  beds, there would be on average 2 alternative beds in use.



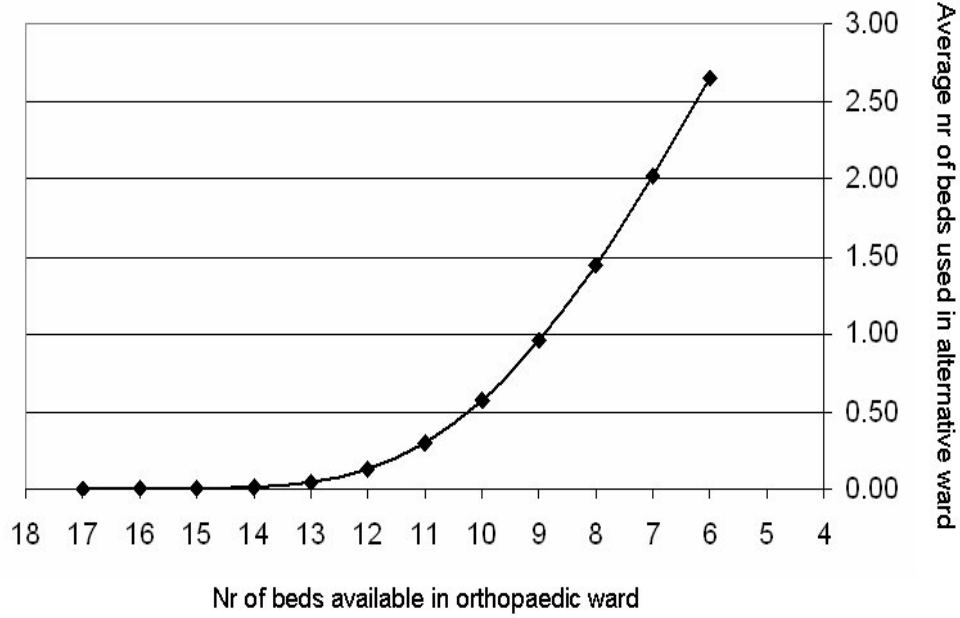


Figure 4: Impact of restricting the number of beds in the orthopaedic ward on the alternative ward.

## 4.2 Case B: Cardiac Catheterization Facility

The cardiac catheterization facility of the university hospital Gasthuisberg in Leuven, which is part of the Universitaire Ziekenhuizen Leuven (8250 employees), will be the subject of this second case study. In particular, we will focus on the surgery suite of this facility in which yearly about 6900 surgeries are performed. In order to capture the diversity of surgery types, the model distinguishes between 15 pathways (e.g. coronarography, ICD or PTCA) and incorporates 10 surgeon groups.

One major concern in organizing the cardiac catheterization facility is how to anticipate emergency cases. The unpredictable nature of these emergencies, both in arrival time and duration of the surgery, causes disruptions in the surgery schedule and hence forces the surgeons to perform surgeries in overtime. Today, about 19 percent of all cardiac patients enter the hospital through the emergency entrance. However, about 20 percent of these patients cannot truly be categorized to be emergency cases since they use the emergency service as an alternative circuit in order to speed up their treatment delivery time. This implies that we have to distinguish between *real* and *feigned* emergency cases. Although we cannot prohibit the occurrence of real

emergency cases, we can try to tackle the wish of a patient to act as a feigned emergency case. In other words, we have to adjust the organization of the cardiac catheterization facility so that the feigned emergency cases want to enter the hospital as regular and thus planned patients.

In a first alternative (B1), we will proportionally adjust the arrival rate of both the regular patients (increase) and the emergency patients (decrease). The feigned emergencies will now enter the hospital as regular patients. The switch results, though, in an unstable simulation system since the work in progress (WIP) is continually increasing due to a lack of surgical capacity. This result could somehow be expected, otherwise there would be no incentive for patients to use the emergency circuit. In Figure 5, the WIP is depicted for each surgeon group. We notice a strong upward trend, and hence capacity problems, for surgeon groups 10, 5, 6, 4 and 2. Note that we cannot rightfully interpret performance indicators when the system is unstable, so that organizational modifications are inevitable.

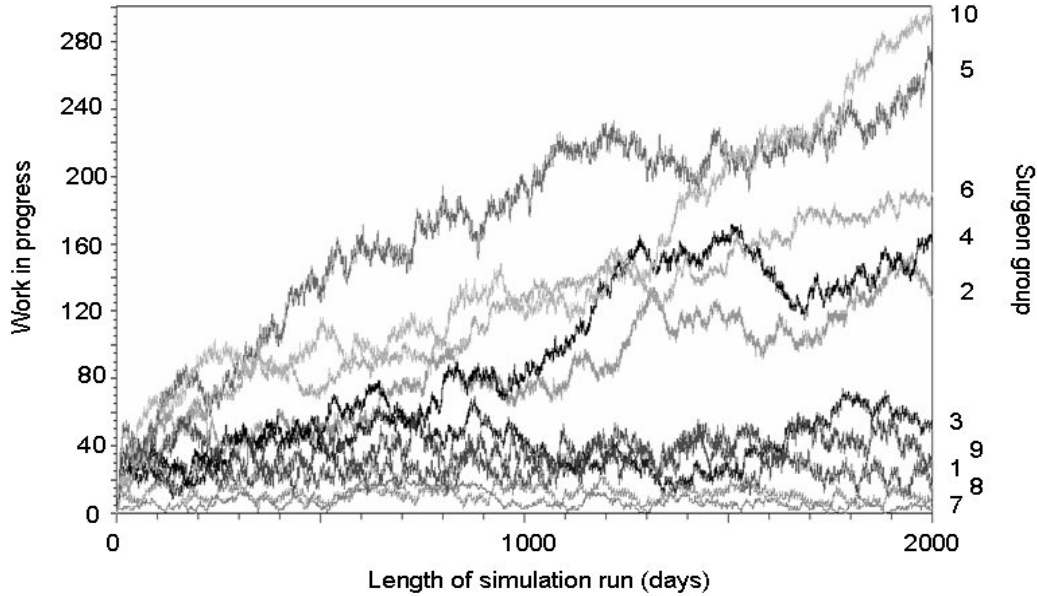


Figure 5: Visualizing the work in progress over time (days) for the surgeon groups.

Two further adjustments were made to alternative B1 in order to stabilize the WIP. First, we tried to adapt the surgery schedule to the increased patient flow. Note, however, that this is a difficult task in reality since the surgery schedule strongly interferes with other schedules (e.g. consultation schedules) or other facilities. Moreover, there might be an intense dislike of changing the working habit. Therefore, as can be illustratively seen from Figure 6, we will only apply

minor adjustments to the surgery schedule. Suppose, in the original surgery schedule (Figure 6 a), that two surgeon groups share one operating room. The first group performs surgeries from 8.00h till 12.00h, whereas the second group is scheduled to be active between 13.00h and 16.30h. The surgery sessions are separated by a one hour break. Suppose, furthermore, that capacity for the first surgeon group has to be expanded. On the one hand, we could eliminate the break and assign that hour to the first surgeon group (Figure 6 b). On the other hand, we could also schedule this surgeon group to start one hour in advance, i.e. at 7.00h (Figure 6 c). Obviously, combinations of both adjustment types are allowed. Note that when the capacity for the second surgeon group had to be augmented, both a break elimination or a late finish would be possible. A second adjustment, made in order to stabilize the WIP, entails a slight modification in the proportion of feigned emergency cases assigned to each surgeon group. This way we want to diminish spare capacity of surgeon groups that are not characterized by an increasing WIP.

The impact of this first alternative on some interesting performance indicators is summarized in Table 3. Remark that the rescheduled surgeries point at patients for which the surgery was postponed until a prioritized emergency surgery was finished. When the shift of a surgeon finishes at the time an emergency surgery is performed in the corresponding operating room, no overtime will be initiated and the remaining, planned surgeries will be canceled. We notice a significant decrease with respect to the current facility practice (BASE) in both the rescheduled and canceled surgeries due to the decreased arrival rate of the emergency cases. In other words, the planned surgical program is less interrupted and hence delayed by the emergencies. Although we did expect a similar decrease in the deviation from the planned surgery start time and the percentage of surgeries started in overtime, the results are contradictory. This implies

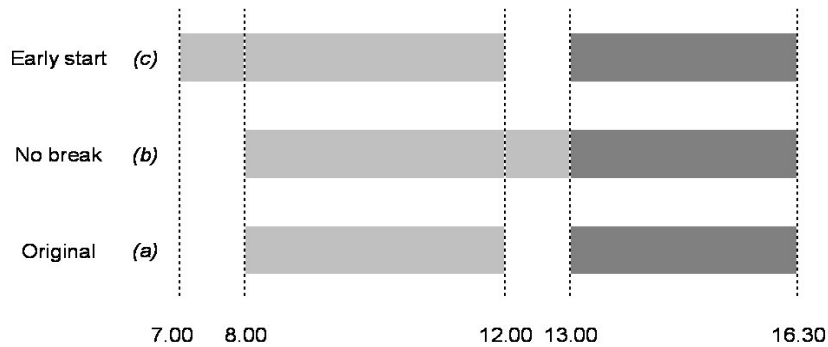


Figure 6: Alternatives in order to modify the surgery schedule.

Table 3: Performance indicators for the tested alternatives (Case B).

	BASE			B1			B2		
Deviation from surgery start time (minutes)	20.59 (0.33)			25.48 (0.31)			18.11 (0.15)		
	Ba	B1	B2	Ba	B1	B2	Ba	B1	B2
	×	✓	✓	✓	×	✓	✓	✓	×
Rescheduled surgeries (%)	11.47 (0.18)			8.22 (0.13)			6.87 (0.08)		
	Ba	B1	B2	Ba	B1	B2	Ba	B1	B2
	×	✓	✓	✓	×	✓	✓	✓	×
Canceled surgeries (%)	1.73 (0.06)			1.02 (0.10)			0.04 (0.01)		
	Ba	B1	B2	Ba	B1	B2	Ba	B1	B2
	×	✓	✓	✓	×	✓	✓	✓	×
Surgeries started in overtime (%)	3.33 (0.08)			3.25 (0.08)			1.32 (0.05)		
	Ba	B1	B2	Ba	B1	B2	Ba	B1	B2
	×		✓		×	✓	✓	✓	×
Late day-care dismissals (%)	10.00 (0.15)			9.29 (0.22)			1.52 (0.09)		
	Ba	B1	B2	Ba	B1	B2	Ba	B1	B2
	×	✓	✓	✓	×	✓	✓	✓	×

✓ denotes a significant difference at a 95 percent confidence level

(x.xx) denotes the half width for a 95 percent confidence interval

that the beneficial effect of the reduction in emergency cases is neutralized or even surpassed by a negative effect, mainly caused by the break elimination in the adapted surgery schedule. In this adapted surgery schedule, several breaks are replaced by planned surgeries, which actually boils down to a buffer removal and hence a reduced probability to get back on schedule when delays occur. The decrease in the late day-care dismissals is barely significant for a similar reason. The beneficial impact of a reduction in emergency cases is almost neutralized by the negative effect of break eliminations and thus enlarged operating room availability. Since day-care patients can now be scheduled later on the surgery day, there is an increased probability to leave the hospital after a certain dismissal limit.

In a second alternative (B2), we will elaborate on alternative B1 and we will introduce three

further modifications. Today, emergency cases are treated in operating room 1 up to 3. This results in an unbalanced spread of the inconveniences caused by the emergencies over the different surgeon groups. So, in alternative B2, we will allow that emergency cases are also assigned to the fourth operating room. The fifth operating room is not available for emergencies since this room is almost exclusively used for children, which inherently already have a high priority. Next, we will also prohibit surgeons to have unexpected and unscheduled breaks during their surgery shift. Such breaks, for instance triggered by the reception of representatives or agents, cause delays and should hence be avoided. Finally, day-care patients will be scheduled near the start of a surgery shift and should precede hospitalized patients. In other words, we will structure the miscellaneous sequence in which surgeries were performed.

When we take Table 3 at hand again in order to evaluate alternative B2, we notice a significant improvement in all the listed performance indicators, both with respect to BASE and alternative B1. By means of illustration we will show that the contribution of sequencing the surgeries is not limited to the decrease in the percentage of late day-care dismissals. This feature could also indirectly influence other performance indicators. In Figure 7 (a), four surgeries have to be performed between 8.00h and 13.00h. Except for surgery 1, all surgeries represent day-care patients. Typically, day-care interventions are highly standardized (i.e. a smaller variance) and have a limited duration. When priority has to be given to an emergency case, both surgery 2 and 3 have to be started in overtime and additional delays w.r.t. the planned surgery duration occur for three patients, namely 2, 3 and 4. When an alternative sequence would be applied, as represented in Figure 7 (b), only one patient will suffer a delay. Moreover, no surgeries have to be started in overtime. It should be clear, though, that the absence of emergency cases would increase the performance of the cardiac catheterization facility as a whole.

## 5 Conclusions and Future Research

In this research paper, a discrete-event simulation model was introduced in order to evaluate issues related to the capacity and the patient flow of clinical pathways. The standardized nature of these pathways contributes to the structure embedded in the simulation model and augments the applicability of this medical operations research tool. Many options can be specified in order to truthfully represent both the consultation and the surgery suite and consequently incorporate

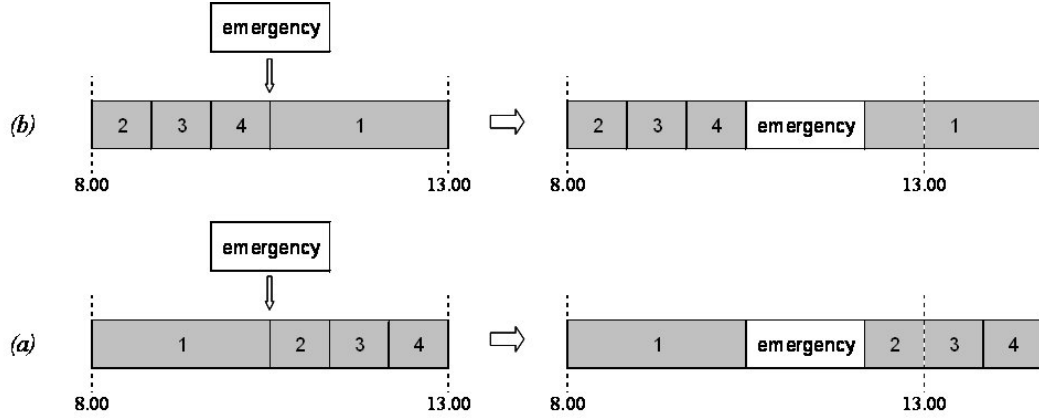


Figure 7: Visualizing the impact of surgery sequencing.

the requested degree of detail. The ability of the simulation model to perform a variety of sensitivity analyses was illustrated by means of two case studies. Although both case studies belong to different medical specialties, they are conceptually similar, i.e. they can both be represented by a combination of clinical pathways and hence investigated by the generic simulation model.

During the conduct of the case studies, difficulties were encountered in modifying the matrix data structures. Although these matrices prevent model users to modify the modules and consequently protects the simulation model for inappropriate use, they clearly make data importation an abstract process. In other words, full attention or profound experience of the model user is required during the setup of the simulation analysis. This implies that, except for the model builders, the accessibility of the model has to be augmented. The development of a simple user interface will consequently constitute the main focus of future research. This interface should not only deal with the importation of data, but should also focus on generating a simplified and structured report of the relevant output statistics.

## Acknowledgements

We would like to thank Prof. Dr. Walter Sermeus of the Center for Health Services and Nursing Research of the Katholieke Universiteit Leuven for introducing the concept of clinical pathways to our research group. We are grateful to Dr. Karl Brabants and Wim Van de Waeter of the Middelheim hospital in Antwerp, both for contributing their clinical experience when the simulation model was built and for providing the data for case A. We also want to thank Gert

Peeters and Marc Vanesch of the university hospital Gasthuisberg in Leuven for their cooperation in providing the data for case B. We appreciate the contribution of Lode Van Damme to the revision and the testing of the simulation model. We acknowledge the support given to this project by the Bijzonder Onderzoeksfonds of the Katholieke Universiteit Leuven.

## References

- Banks, J., Carson, J., Nelson, B. and Nicol, D. 2001. *Discrete-Event System Simulation*, Prentice Hall, Upper Saddle River, NJ.
- Bowers, J. and Mould, G. 2004. Managing uncertainty in orthopaedic trauma theatres, *European Journal of Operational Research* **154**: 599–608.
- Bowers, J. and Mould, G. 2005. Ambulatory care and orthopaedic capacity planning, *Health Care Management Science* **8**: 41–47.
- Cayirli, T. and Veral, E. 2003. Outpatient scheduling in health care: A review of literature, *Production and Operations Management* **12**: 519–549.
- Couchman, A., Jones, D. and Griffiths, K. 2002. Predicting the future performance of a clinical biochemistry laboratory by computer simulation, *Simulation Modelling Practice and Theory* **10**: 473–495.
- De Bleser, L., Depreitere, R., De Waele, K., Vanhaecht, K., Vlayen, J. and Sermeus, W. 2006. Defining pathways, *Journal of Nursing Management* **14**: 553–563.
- Harper, P. and Gamlin, H. 2003. Reduced outpatient waiting times with improved appointment scheduling: A simulation approach, *OR Spectrum* **25**: 207–222.
- Jun, J., Jacobson, S. and Swisher, J. 1999. Application of discrete-event simulation in health care clinics: A survey, *Journal of the Operational Research Society* **50**: 109–123.
- Kelton, W., Sadowski, R. and Sturrock, D. 2004. *Simulation with Arena*, McGraw-Hill, New York, NY.
- Kim, S.-C., Horowitz, I., Young, K. and Buckley, T. 2000. Flexible bed allocation and performance in the intensive care unit, *Journal of Operations Management* **18**: 427–443.

- Klassen, K. and Rohleder, T. 1996. Scheduling outpatient appointments in a dynamic environment, *Journal of Operations Management* **14**: 83–101.
- Ridge, J., Jones, S., Nielsen, M. and Shahani, A. 1998. Capacity planning for intensive care units, *European Journal of Operational Research* **105**: 346–355.
- Rohleder, T. and Klassen, K. 2000. Using client-variance information to improve dynamic appointment scheduling performance, *Omega - The International Journal of Management Science* **28**: 293–302.
- Swisher, J., Jacobson, S., Jun, J. and Balci, O. 2001. Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation, *Computers & operations research* **28**: 105–125.
- Visser, J., Adan, I. and Dellaert, N. 2006. Developing a platform for comparison of hospital admission systems: An illustration, *To appear in European Journal of Operational Research*.
- Yeung, R., Leung, G., S., M. and Johnston, J. 2004. Waiting time and doctor shopping in a mixed medical economy, *Health economics* **13**: 1137–1144.
- Zander, K. 2002. Integrated care pathways: Eleven international trends, *Journal of Integrated Care Pathways* **6**: 101–107.